

Available online at www.sciencedirect.com**ScienceDirect**

Procedia - Social and Behavioral Sciences 95 (2013) 170 – 177

Procedia
Social and Behavioral Sciences

5th International Conference on Corpus Linguistics (CILC2013)

A Corpus-based Study on Referential Chains in Spanish Newspapers

An Vande Castele*

Vrije Universiteit Brussel, Faculteit Letteren & Wijsbegeerte, Pleinlaan 2, 1050 Brussel, Belgium

Abstract

This article clarifies the choice of referential devices in Spanish newspapers. First, there is a wide range of referential devices which can be divided into two major types: full lexical items (full lexical noun phrases, proper names) and reduced items (such as pronouns and verbal morphemes). The selection of an appropriate linguistic expression in a discourse is subject to a complex process which depends on different semantic, syntactic and pragmatic elements. This paper describes how syntax relates to semantic content and pragmatic function in Spanish newspaper discourse.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](#).
Selection and peer-review under responsibility of CILC2013.

Keywords: Reference ; Accessibility ; Anaphora ; Coreference ; Newspaper Discourse

1. Introduction

One of the more interesting facts about human language is that we can use different forms to refer to the same thing, and the same form can be used to refer to many different things. Yet people somehow manage to understand one another (Gundel, Hedberg and Zacharski, 1993: 274).

Referential management is necessary to be able to create utterances which can be clearly understood. This means that referential devices are chosen in function of the discourse role they assume. So, there is an important correlation between form or referent and meaning or reference and it is the degree of mental accessibility of a referent which

* Corresponding author. Tel.: +32-2-629-26-51; fax: +32-2-629-36-84.
E-mail address: An.Vande.Castele@vub.ac.be

determines how reference should be realized (Ariel, 1990; Givon, 1983; Gundel, Hedberg and Zacharski, 1993; Prince, 1985).

The exploratory study presented in this article intends to describe which factors govern the choice of a particular linguistic expression in Spanish newspaper discourse. The present analysis further builds upon previous research (*cf.* Vande Castele, 2010a; 2010b), but gives new insights by enlarging the types of analyzed data. While our previous investigations focused on full lexical noun phrases and proper names, now all kinds of referring expressions are taken into account. This means that also reduced referential devices such as pronouns and verbal morphemes are examined.

2. Referential choice as a multi-factorial process

Referential choice needs to be considered as a “*linguistic process of a multi-factorial nature*”, argues Kibrik (2011: 61) and the range of parameters determining referential status is broad. According to Grüning and Kibrik (2005) these parameters can be grouped in two main classes. Some properties belong to the referred entity, such as animacy and centrality in the discourse. Others are discourse related like the distance to the previously mentioned antecedent, the cognitive status of the referent, possible competitors in the discourse, etc.

Dooley and Levinsohn (2001: 112) point out that a speaker completes several tasks when choosing a particular linguistic expression: a semantic task in order to identify referents unambiguously by distinguishing them clearly from other possible referents and a discourse-pragmatic task to signal the informative status, the saliency and the prominence of a referent. Finally the information processing serves to avoid any disruption in the information flow. These ideas also respond to Grices (1975) maxim of quantity in the sense that each contribution should be as informative as required but also not more informative than necessary. So, the choice for a particular linguistic device is not random.

Ariel (1990) proposed a graded scale to account for the cognitive status of referents. The basic idea is that accessibility is considered as a continuum. So entities are accessible to a greater or lesser degree. Now, Ariel shows in her scale that accessibility status correlates to the form of the referent used.

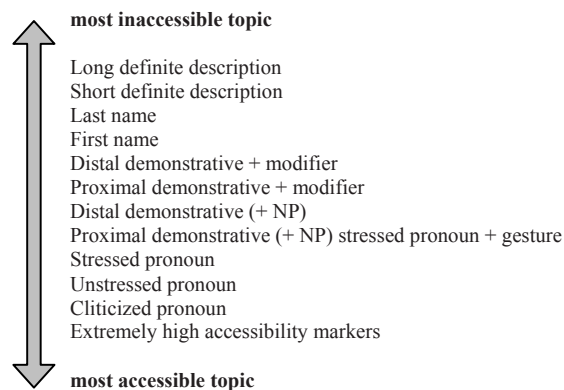


Fig. 1. Ariel's accessibility scale (1990: 73)

Longer linguistic forms, such as full lexical noun phrases tend to be used for low accessible referents and shorter forms seem more appropriate for highly accessible items.

The more informative, rigid, and unattenuated an expression is, the lower the degree of accessibility it codes, and vice versa, the less informative and rigid and the more attenuated the form is, the higher the accessibility it codes.” (Ariel, 2001:32)

Ariel also asserts that the following four factors are essential in affecting the processing of referential expressions:

- Distance: the distance between the antecedent and the anaphor
- Competition: the availability of potential alternative competitive antecedents
- Saliency: the degree of topicality or discourse prominence
- Unity: the unity of the frames in which the anaphor and the antecedent appear

In this sense, the informational degree of an expression can be quantitatively described, since the selection of a referent depends on the quantity of information considered as necessary in order to enable the referential process.

3. Corpus-based analysis of referents

As stated before, the aim of the present study on entire referential chains is to provide insights on the most frequently used referential expressions in informative articles published in Spanish newspapers. Whereas previous work (Vande Castele, 2009; 2010a; 2010b) was centered around different kinds of nominal expressions, this analysis broadens the perspective adding also pronominal and zero anaphora. So, generally speaking the objective is to find out how syntactic structures can be related to a particular semantic content and a pragmatic function.

For the study a corpus of Spanish newspaper articles has been compiled in which human referents were annotated. The texts are informative articles published in the newspapers *ABC* and *El País* in November and December of 2012. It should be observed that we compiled our own corpus, because of the purpose of our study. Since we want to investigate complete referential chains, it is important to have the articles in their totality and therefore we could not work with existing corpora such as *CREA*, the Reference Corpus of the Real Academia Española or the *Corpus del Español*, provided by Mark Davies.

A database of 295 referential expressions has been compiled. These expressions are used to refer to a hundred different persons. This means that the mean would be 3 references to each entity, but we have to underline that the below figures show a large variety, going from one expression per person (in 43 cases) to 14 consecutive references (for instance to indicate the German chancellor Angela Merkel).

The next table shows how many references were counted for each referent.

Table 1. Number of references per referent.

One reference	43
2 consecutive references	17
3 consecutive references	17
4 consecutive references	7
5 consecutive references	3
6 consecutive references	4
7 consecutive references	4
9 consecutive references	1
11 consecutive references	1
12 consecutive references	1
13 consecutive references	2
14 consecutive references	1
Total	295

The next paragraph details on how the data were analyzed. First all different kinds of linguistic devices referring to human beings were annotated: proper names, appositive constructions, pure descriptive noun phrases, pronouns, etc. The following example (*ABC*, 03/12/12) serves as an illustration of how we proceeded:

Alemania dice que la recapitalización directa no llegará a tiempo a España

EUROPA PRESS / BRUSELAS

ABC 03/12/2012 - 17.25h

Schäuble asegura que nuestro país sufre por el contagio de la crisis de la eurozona y no tiene otros problemas aparte del bancario

El ministro alemán de Finanzas, Wolfgang Schäuble, ha insistido este lunes en que la recapitalización directa de la banca a cargo del fondo de rescate (MEDE) tardará en ponerse en marcha y no llegará a tiempo para el caso de España. "España, por supuesto, no puede esperar tanto" para la recapitalización directa, ha afirmado **Schäuble** en una comparecencia conjunta con **su homólogo francés, Pierre Moscovici**, ante la comisión de Asuntos Económicos de la Eurocámara. "Espero que en el Eurogrupo completemos (el rescate bancario) más pronto y que la capitalización bancaria sea posible utilizando los medios del MEDE pero yendo a través del Estado español", ha apuntado **el ministro alemán de Finanzas**. "Decir que los bancos tendrán acceso inmediato directo al MEDE es una expectativa exagerada", **ha resaltado**.

appositive construction

proper name

descriptive noun phrase

morpheme

Primero, **ha recordado**, es necesario aprobar el supervisor bancario único, algo que intentará lograr el Ecofin este martes.

Fig. 2. Illustration of the examined linguistic expressions

The table below indicates the frequency of each kind of linguistic expression in our database.

Table 2. Distribution of referents.

Appositive constructions	99
Proper names	109
Descriptive noun phrases	52
Morphemes	28
Pronouns	1
Possessive determiner	4
Copular constructions	2
Total	295

For each expression, the morphosyntactic properties were examined and the informative role it assumes in the text was defined. The analysis has been conducted in a systematic way, starting with the first reference to the entity in the text and then determining all anaphoric relations to the entity.

In a previous article, presented at the first CILC-conference (Vande Castele, 2009), the role of references in headlines were detailed. We then stated that headlines and subtitles have a particular role in discourse: more than the informative aspects, they focus on attracting reader's attention. Therefore, in the present study, references in headlines and subtitles were not taken into account. So the first reference to an item appears really in the body of text.

A detailed analysis of the first references to persons in the newspaper articles shows that in most cases a proper name is joined to a descriptive part. These structures are called appositive constructions and can be defined in the following way. First, some examples:

- **El ministro de Asuntos Exteriores y Cooperación, José Manuel García-Margallo**, advirtió ayer en Valencia de la urgencia de resolver los problemas de terrorismo integrista que afectan a la zona norte de África, ya que de lo contrario España podría acabar siendo «frontera» de Al Qaeda. (*ABC*, 04/12/12)
- **El presidente de Egipto, Mohamed Mursi**, *llegará a un acuerdo con el Consejo Supremo Judicial*, el máximo órgano de gobierno del poder judicial egipcio, sobre la propuesta de limitar el alcance del decreto presidencial publicado el pasado jueves que permitiría la ampliación de los poderes del gobernante, según ha declarado este lunes **el ministro de Justicia, Mekky Ahmed**. (*ABC*, 27/11/12)

The appositive constructions start with a descriptive part in which the public role of the person is presented and then a denominative part follows with the person's name. These are expressions with an important informative charge. In Vande Castele (2010a) we called these expression specificative appositive constructions. They are so-called loose appositions (Acuña Fariña, 1996; Burton-Roberts, 1975; Forsgren, 2000; Martínez, 1994; Meyer, 1992; Neveu, 1998; Suñer Gratacós, 1999; etc.) in which the second segment identifies the reference described in the first part. In our database, 88 of the 101 first references are appositive constructions. In the next table, the results presented previously in table 2, are detailed for the first and consecutive references.

Table 3. Detailed distribution of referents.

	First reference	Consecutive references	Total
Appositive constructions	88	11	99
Proper names	9	100	109
Descriptive noun phrases	2	50	52
Morphemes		28	28
Pronouns		1	1
Possessive determiner		4	4
Copular constructions	2		2
Total	101	294	295

Within this group of 88 appositive constructions, used to introduce a new referent in the newspaper article, the vast majority are specificative appositive constructions as described and illustrated above. So they first state the public role of an individual and then specify their name. Only 7 show another internal structure. Their segments appear in the opposite order, since the proper name appears first and is then detailed in a descriptive part. The informative role of these structures is more descriptive rather than identificative.

Our results also clearly indicate that full lexical devices are more often used to refer to already introduced individuals. Proper names and descriptive noun phrases appear respectively in 100 and 50 cases, while reduced items only appear in 33 instances (28 morphemes, 1 pronoun and 4 possessive determiners).

The following paragraphs cited from our corpus data serve to clarify the results.

- **La canciller alemana Angela Merkel** criticó el jueves la política de asentamientos israelí ante **el primer ministro Benjamín Netanyahu**. **El israelí** se había reunido en Berlín con **su homóloga alemana** para la cumbre anual entre los Ejecutivos de ambos países. La víspera, **Netanyahu** se había dicho “decepcionado” por la abstención de Alemania en la votación en la que Naciones Unidas elevó el estatus de Palestina a “Estado observador”. Durante la rueda de prensa que ofreció junto a **Netanyahu** este mediodía, **Merkel** defendió el “derecho de Israel a la autodefensa” y recordó que la seguridad y la existencia de Israel, “la única democracia” en la región de Oriente Próximo, es “razón de Estado” para Alemania. **La democristiana Merkel (CDU)** insistió

no obstante varias veces en sus críticas a la política de asentamientos de Israel en los territorios palestinos ocupados. Para la primera de ellas, **la canciller** se giró con una sonrisa hacia **Netanyahu**: “Estamos de acuerdo en que no estamos de acuerdo”. Preguntada sobre la posibilidad de dar “más pasos” para disuadir a Israel de estas políticas, **Merkel** dijo que ella no es “una persona que amenace” y recalcó que el compromiso alemán con Israel “no se debe únicamente a una horrible experiencia histórica”. En este sentido, expresó la “alegría alemana por poder cooperar de forma tan estrecha con Israel” siete décadas después de la Shoá. **Merkel** anunció un nuevo estrechamiento de las relaciones entre Alemania e Israel con ocasión del 50º aniversario del establecimiento de relaciones diplomáticas entre ambos países, en 1965. [...] (*El País*, 06/12/12)

References to the German chancellor Angela Merkel are made by the means of an appositive construction “*la canceller alemana Angela Merkel*”, descriptive noun phrases “*su homóloga alemana*” and “*la canceller*”, the proper name “*Merkel*” which appears several times; Only one reference is realized by a verbal morpheme in “*expresó*”. “*La democristiana Merkel (CDU)*” is analyzed as a so-called close apposition. These one membered appositions are used for further references.

In the references to the Prime Minister of Israel less variety is noted. We observe one appositive construction to introduce the referent. Further mentions are mostly realized by his name and only once in a descriptive way “*el israelí*”.

In the next example three protagonist appear and alternate often. Therefore lexicalized references are required. It is obvious that morphemes or pronouns appear less, since they do not offer enough information to avoid ambiguity.

- **La canciller de Alemania, Angela Merkel**, ha afirmado este viernes que pese a la falta de acuerdo en la cumbre extraordinaria de presupuesto de la Unión Europea (UE), hay “suficiente potencial” para que los países alcancen un pacto a principios de año. “Creo que es posible alcanzar un acuerdo entre los Veintisiete a principios de año”, ha dicho **Merkel** durante una conferencia de prensa al término de la reunión de dos días en la que han quedado escenificadas las diferencias de los líderes sobre la dotación presupuestaria de la UE para el periodo 2014-2020. **La canceller alemana** ha asegurado que **el presidente del Consejo Europeo, el belga Herman Van Rompuy**, “seguirá trabajando para tratar de encontrar el consenso en las próximas semanas”. Por su parte, **el presidente de Francia, François Hollande** dijo hoy que el volumen de gasto global planteado para el presupuesto por **Van Rompuy**, es “razonable” y confió en que será posible lograr un acuerdo entre los Estados miembros a principios de 2013. **Hollande**, en todo caso, se mostró contrario a aceptar una reducción importante sobre esa cifra, que ya está por debajo del billón de euros para el periodo 2014-2020, oponiéndose claramente a lo que pide un grupo de países encabezado por el Reino Unido. “Queremos un presupuesto consistente, que tenga posibilidad de financiar las políticas europeas. Lo propuesto por **Van Rompuy** nos parece razonable”, dijo **el presidente galo**, que subrayó que la reunión de los Veintisiete ha sido “útil” y supone “una etapa” de cara a lograr un consenso. (*ABC*, 23/11/12)

Nevertheless, even in cases where only one referent is maintained during a certain span in the text, proper names and short lexical noun phrases such as “el president” or “el ministro” appear, as exemplified by the references to the British Prime Minister, David Cameron:

- **El primer ministro David Cameron** respalda la propuesta, que puede llevarle a enfrentamientos con los parlamentarios más conservadores, así como con la Iglesia Anglicana y la Católica, según informa este viernes el diario británico The Guardian. **Cameron** apoya este compromiso a través de lo que se conoce como cláusula permisiva que permite que los matrimonios homosexuales se celebren lugares de culto, pero no obliga a las organizaciones religiosas a celebrarlos. «Soy un gran defensor del matrimonio y no quiero que las personas gays sean excluidas de esta gran institución. Pero seré claro, si no hay ninguna iglesia o sinagoga o mezquita que quiera celebrar una boda homosexual no lo harán, de ninguna manera se verán forzadas a hacerlo», aclaró **Cameron**, que aseguró que esta cuestión está absolutamente clara en la legislación. «La votación de los miembros del parlamento será libre, pero yo personalmente lo apoyaré», añadió **el primer ministro**. (*ABC*, 07/12/12)

Only in a few cases appositive constructions were found to indicate an already introduced referent. As illustrated in the following example, certain discourse factors enhance the use of an appositive construction. First, the distance between the first and the second reference to the Vice-president Diego Valderas is large and there are also a lot of potential competitive antecedents in between the two references.

- La delegada del Gobierno en Andalucía, Carmen Crespo, se sumó este viernes a las advertencias del ministro de Hacienda, Cristóbal Montoro, sobre el cumplimiento de déficit (1,5% del PIB para las comunidades autónomas) e instó a la Junta a que establezca “medidas adicionales” si, como apuntan las previsiones, lo supera. Crespo afirmó que el Ejecutivo está dispuesto a negociar las “fórmulas o métodos” para conseguirlo. Los últimos datos de Hacienda sitúan el déficit andaluz en el 1,6%. “Pasarse unas décimas no es el fin del mundo”, ha replicado **el vicepresidente de la Junta, Diego Valderas (IU)**. Carmen Crespo, que visitó la sede del Ayuntamiento de Adra, instó a la Junta a “autolimitarse” para llegar a sus objetivos sin necesidad de que el Gobierno central haga uso de las medidas coercitivas previstas en la norma que regula el Fondo de Liquidez Autonómico (FLA), que ha permitido a la Junta obtener 2.133 millones, aunque Griñán había reclamado 4.900. “El fin”, ha dicho Crespo, “es que entre todos se dé el empujón hacia la senda del crecimiento y la creación de empleo”. “Lo que tenemos clarísimo desde el Gobierno de España es que todos tenemos que cumplir y poner en orden nuestras cuentas públicas, la deuda y el déficit”, ha explicado Crespo, quien ha advertido que, pese a las medidas previstas en el FLA, “no está vista” una posible intervención de Andalucía. En este sentido, la representante del Gobierno ha insistido en que antes hay que agotar las líneas de negociación abiertas sobre la fórmula de devolución del Fondo de Liquidez Autonómico (FLA), que estaría supeditada a “qué condiciones tiene que cumplir” y “en lo que tiene que hacer la comunidad autónoma para ponerse al día”. **El vicepresidente de la Junta y consejero de Administración y Relaciones Institucionales, Diego Valderas (IU)**, criticó la actitud del Gobierno y calificó de “amenazas” las advertencias y exigencias de Montoro sobre el déficit. (*El País*, 07/12/12)

Finally, one would assume that well know people need less informative devices, but as illustrated by these last examples, even the president of the United States, Obama, and the First Minister of Spain, Rajoy, are specified by the means of an appositive construction in the Spanish newspapers.

- **El presidente del Gobierno, Mariano Rajoy**, ha avisado este martes al de la Generalitat, Artur Mas, de que «nadie va a sacar a los catalanes de España ni de Europa», y ha reclamado un último esfuerzo para que el PP logre unos resultados históricos en las elecciones del 25 de noviembre. (*ABC*, 21/11/2012)
- **El presidente de Estados Unidos, Barack Obama**, alertó hoy a los republicanos de que no tolerará una nueva pelea sobre el techo de la deuda en el marco de las negociaciones para evitar el "precipicio fiscal", estancadas fundamentalmente por el desacuerdo sobre los impuestos a los ricos. (*El País*, 05/12/2012)

4. Corpus-based analysis of referents

As a conclusion, our analysis showed that although pronouns and morphemes seem the most appropriate linguistic expression to refer to an already mentioned entity in a discourse, in our corpus of Spanish newspaper articles they only appear in a few occasions and full lexicalized expressions are preferably used. Sometimes this is due to the fact that newspaper articles are highly informative and the use of reduced devices would lead to ambiguity. In other articles, several protagonists appear in the same paragraph and to only way to clearly indicate the correct antecedent is by repeating the proper name or his public role.

Moreover we noted a high frequency of appositive constructions which combine a proper name with a descriptive segment. So, despite the fact that proper names are defined as rigid designators (Kripke, 1972), which implies that proper names on their own should be able to establish a reference, our investigation clearly shows that proper names alone do not appear to be as informative or newsworthy as required for the discourse purpose. Even well know people are fully introduced by their name and the role they assume in public life.

References

- Acuña Fariña, J. C. (1996). *The Puzzle of Apposition. On So-Called Appositive Structures in English*. Santiago de Compostela: Universidade de Santiago de Compostela.
- Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Burton-Roberts, N. (1975). Nominal Apposition. *Foundations of Language*, 13, 391-419.
- Dooley, R. A. & Levinsohn, S. H. (2001). *Analyzing discourse: A Manual of Basic Concepts*. Dallas: SIL International.
- Forsgren, M. (2000). Apposition, attribut, épithète: même combat prédicatif? *Langue française*, 125, 30-45.
- Givon, T. (1983). Topic Continuity in Discourse: An Introduction. In T. Givon (Ed.), *Topic Continuity in Discourse. A Quantitative Cross-Language Study* (pp. 5-41). Amsterdam: John Benjamins.
- Gundel, J. K., Hedberg, N. & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69, 274-307.
- Kibrik, A. (2011). *Reference in Discourse*. New York: Oxford University Press.
- Kibrik, A. & Grüning A. (2005). Modelling Referential Choice in Discourse: A Cognitive Calculative Approach and a Neural Networks Approach. In A. Branco, T. McEnery & R. Mitkov (Eds.), *Anaphora Processing: Linguistic, Cognitive and Computational Modelling* (pp. 163-198). Amsterdam: Benjamins.
- Kripke, S. A. (1972). Naming and Necessity. In D. Davidson & G. Harman (Eds.), *Semantics of Natural Language* (pp. 253-355). Dordrecht: Reidel.
- Martínez, J. A. (1994). Las construcciones apositivas en español. In J.A. Martínez (Ed.), *Cuestiones marginadas de gramática española* (pp. 173-224). Madrid: Istmo.
- Meyer, C. F. (1992). *Apposition in Contemporary English*. Cambridge: Cambridge University Press.
- Neveu, F., (1998). *Etudes sur l'apposition. Aspects du détachement nominal et adjectival en français contemporain dans un corpus de textes de J.-P. Sartre*. Paris: Honoré Champion.
- Prince, E. F. (1981). Toward a Taxonomy of Given-New Information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-256). New York: Academic Press.
- Suñer Gratacós, A. (1999). La aposición y otras relaciones de predicación en el sintagma nominal. In I. Bosque & V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española. Primera parte. Sintaxis básica de las clases de palabras* (pp. 523-564). Madrid: Espasa Calpe.
- Vande Castele, A. (2009). El papel comunicativo de las expresiones nominales con referente humano en los titulares de artículos periodísticos. In A. Sánchez Gómez & P. Cantos Pérez (Eds.), *A survey of Corpus-based Research* (pp. 1045-1055). Murcia: Asociación Española de Lingüística del Corpus.
- Vande Castele, A. (2010a). *Las construcciones apositivas de referente humano en el discurso periodístico informativo*. Tübingen: Narr Francke Attempto Verlag.
- Vande Castele, A. (2010b). Linguistic Strategies Referring to Politicians in the Spanish Written Press: an Analysis of Appositive Constructions. *Pragmalingüística*, 18, 206 – 222.